Investigating the exponential distribution

Meenakshi Parameshwaran

25 January 2016

Overview

In this project I investigate the exponential distribution in R and compare it with the Central Limit Theorem. I simulate the exponential distribution using the `rexpn()` function and compare the sample mean of the distribution with the theoretical mean of the distribution. I also compare the sample variance with the theoretical variance. I find that with a large number of simulations (1000), the simulated mean and variance approximate the theoretical mean and variance, and that the simulated distributions are approximately normal.

Simulations

The code chunk below simulates the distribution of 1000 averages of 40 exponentials. Note that lambda (the rate) is 0.2 for all of the simulations.

```
# simulating means
set.seed(seed = 1000) # set the seed
lambda = 0.2
mns = NULL
for (i in 1 : 1000) mns = c(mns, mean(rexp(n = 40, rate = lambda)))
```

The code chunk below simulates the distribution of 1000 variances of 40 exponentials.

```
# simulating variances
set.seed(seed = 1) # set the seed
lambda = 0.2
vrs = NULL
for (i in 1 : 1000) vrs = c(vrs, var(rexp(n = 40, rate = lambda)))
```

Sample Mean versus Theoretical Mean

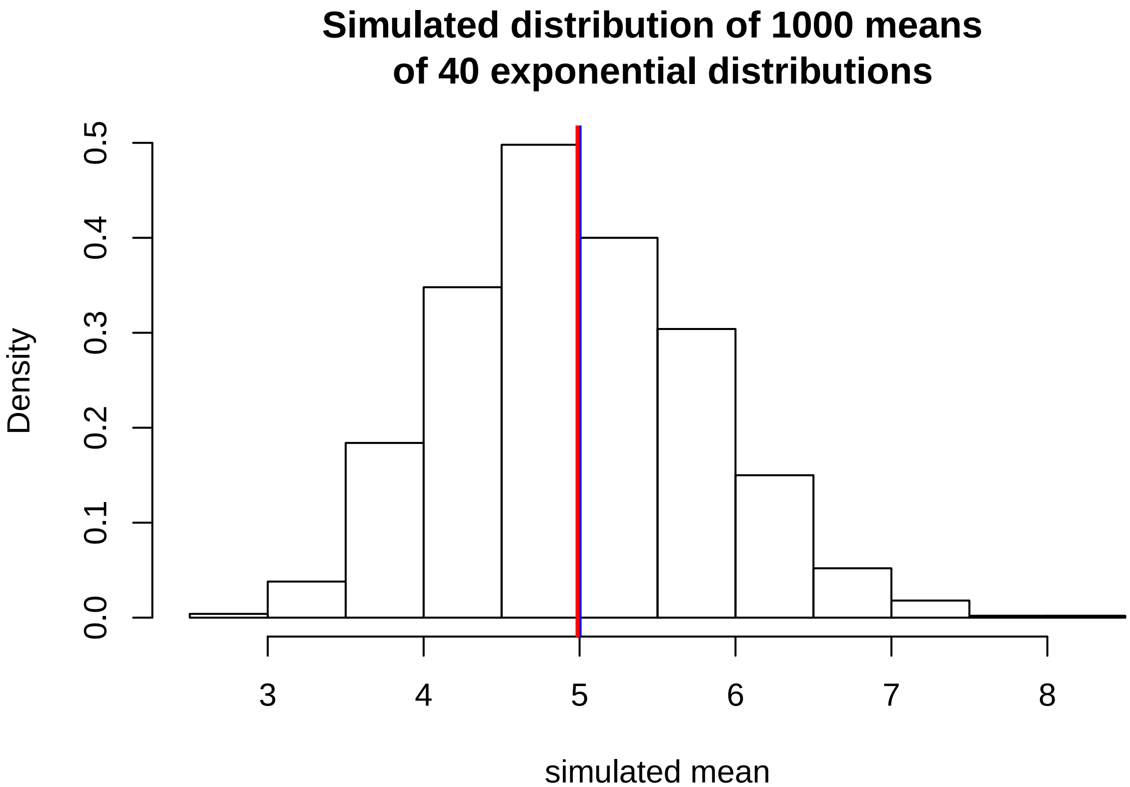
The figure below shows a histogram of the simulated distribution of 1000 means of 40 exponential distributions. The theoretical mean of the exponential distribution is `1/lambda`, which is 5 (blue line). The sample mean of the simulated distribution is `mean(mns)` which is 4.9869634 (red line).

Because the simulated mean is very close to the theoretical mean, the red and blue lines are essentially overlapping. This indicates that the simulated mean is a good approximation of the theoretical mean of the exponential distribution.

```
# histogram
hist(mns, xlab = "simulated mean", main = "Simulated distribution of 1000 means \n of 40 exponential di
    freq = FALSE)

# add theoretical mean line
abline(v = 1/lambda, col = "blue", lwd = 2)
```

```r
# add simulated mean line
abline(v = mean(mns), col = "red", lwd = 2)
```



Sample Variance versus Theoretical Variance

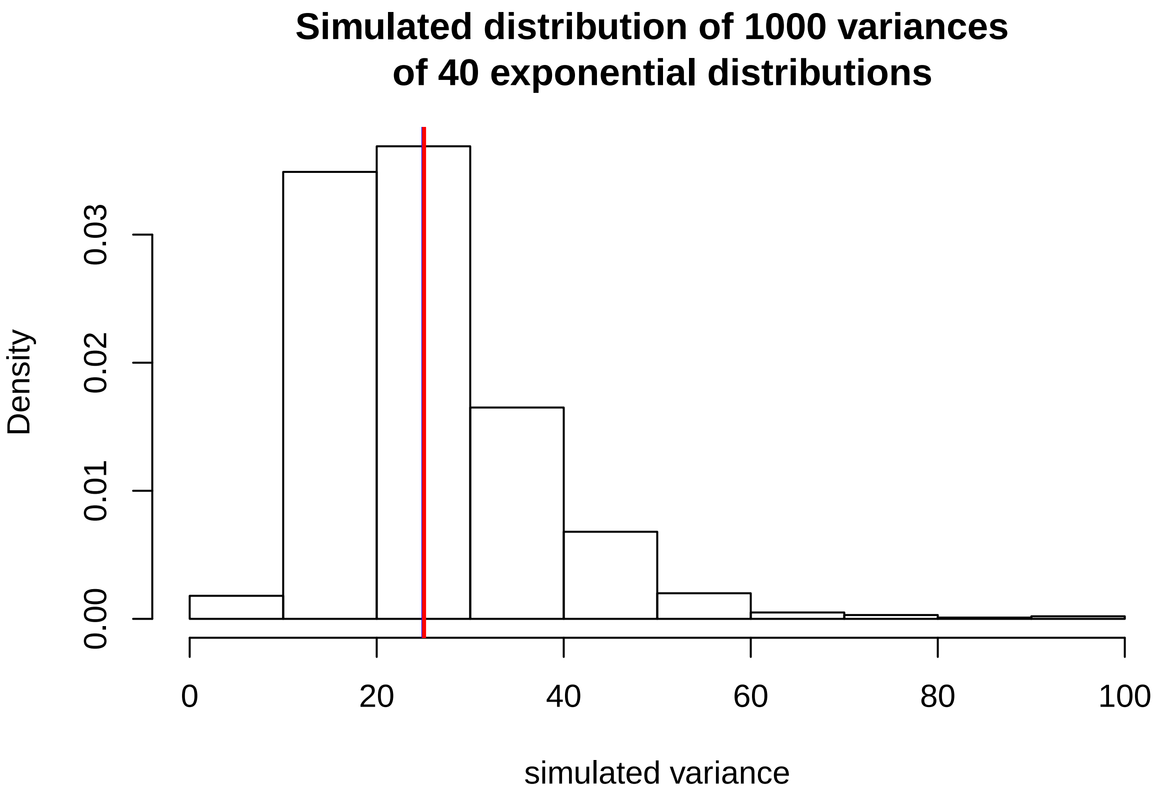
The figure below shows a histogram of the simulated distribution of 1000 variances of 40 exponential distributions. The theoretical variance of the exponential distribution is the standard deviation squared i.e. `(1/lambda)^2`, which is 25 (blue line). The mean sample variance of the simulated distribution is `mean(vrs)` which is 25.064586 (red line).

Because the simulated mean sample variance is very close to the theoretical variance, the red and blue lines are almost perfectly overlapping. This indicates that the sample variance is a good approximation of the theoretical variance of the exponential distribution.

```r
# histogram
hist(vrs, xlab = "simulated variance", main = "Simulated distribution of 1000 variances \n of 40 expone
    freq = FALSE)

# add theoretical variance line
abline(v = (1/lambda)^2, col = "blue", lwd = 2)

# add simulated variance line
abline(v = mean(vrs), col = "red", lwd = 2)
```



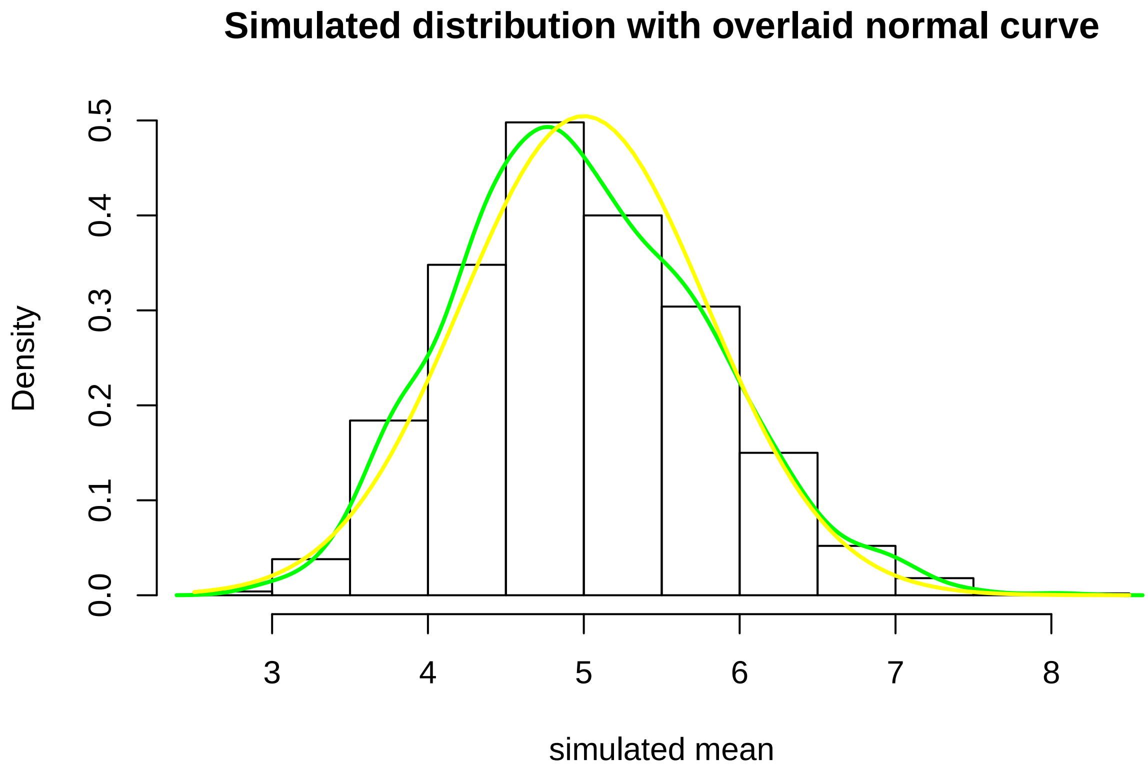
Distribution

The figure below is the same as the first figure above except it has been overlaid with a density plot of the simulated exponential distribution (yellow curve) and with the normal distribution (green curve). It is clear that the simulated distribution closely resembles the normal (Gaussian) distribution, as the Central Limit Theorem would predict (with large N, as in this case - 1000 simulations). This means that the simulated distribution can be considered approximately normal.

```
# histogram
hist(mns, xlab = "simulated mean",
     main = "Simulated distribution with overlaid normal curve",
     freq = FALSE)

# overlay density plot of simulated means
lines(density(mns), col = "green", lwd = 2)

# overlay true normal density plot
x <- seq(from = min(mns), to = max(mns), by = 0.01) # create x values
curve(dnorm(x = x, mean = 1/lambda, sd = (1/lambda)/sqrt(40)), add = TRUE, col = "yellow", lwd = 2) # p
```


Simulated distribution with overlaid normal curve
Density
0.0
0.1
0.2
0.3
0.4
0.5
3
4
5
6
7
8
simulated mean